

Value at risk and the diversification dogma

Arturo Erdely*

Facultad de Estudios Superiores Acatlán
Universidad Nacional Autónoma de México
arturo.erdely@comunidad.unam.mx

Abstract

The so-called risk diversification principle is analyzed, showing that its convenience depends on individual characteristics of the risks involved and the dependence relationship among them.

Keywords: value at risk, loss aggregation, comonotonicity, diversification.

1 Introduction

A popular proverb states *don't put all your eggs in one basket* and it is implicitly based on a *principle* (let's call it that way momentarily) of *risk diversification* which could have the following “justification”: Suppose it is needed to take $2n$ eggs from point A to point B, walking distance, and that there are only two alternatives available, either one person carrying all the eggs in one basket, or two people with n eggs each in separate (and independent) baskets. The proverb suggests that there is a higher risk with the single person alternative since if he/she happens to stumble and fall we would have a total loss, while with the second alternative only half of the eggs would be lost, and in a worst case scenario (with lower probability) where the two people fall the loss would be the same as in the first alternative, anyway.

Let X be a random variable which counts how many eggs are lost under the first alternative (one basket), and let Y account for the same but for the second alternative (two baskets). Let $0 < \theta < 1$ be the probability of falling and breaking the eggs in a basket while walking from point A to B. Then X and Y are discrete random variables such that $\mathbb{P}(X \in \{0, 2n\}) = 1$ and $\mathbb{P}(Y \in \{0, n, 2n\}) = 1$, with point probabilities $\mathbb{P}(X = 2n) = \theta$, $\mathbb{P}(X = 0) = 1 - \theta$, $\mathbb{P}(Y = 2n) = \theta^2$, $\mathbb{P}(Y = n) = 2\theta(1 - \theta)$, and $\mathbb{P}(Y = 0) = (1 - \theta)^2$. Certainly the probability of facing the maximum loss of $2n$ eggs has a higher probability under the first alternative, but it is also true that the *no loss* probability is also higher under such alternative. Moreover:

$$\mathbb{P}(Y > 0) = \theta^2 + 2\theta(1 - \theta) = \theta(2 - \theta) > \theta = \mathbb{P}(X > 0),$$

which means that there is a higher probability of suffering a (partial or total) loss under the second alternative. Therefore... does it mean that it is better to put all the eggs in one basket? If a single trip is going to take place, the answer would be yes, but if the same trip is going to be repeated a large number of times we should analyze the *long run* average loss, which would be $\mathbb{E}(X) = 2n\theta$ for the first alternative, and $\mathbb{E}(Y) = 2n\theta^2 + 2n\theta(1 - \theta) = 2n\theta$ for the second alternative, that is, in the long run there is no difference between the two alternatives.

Is it never more convenient to diversify in two baskets? If the probability of stumbling and falling with $2n$ eggs is the same as with half of them (which might be true up to certain value of n) then the proverb is certainly wrong, but maybe for a sufficiently large value of n we should consider different probabilities of

*Personal website <https://sites.google.com/site/arturoerdely>

falling and breaking the eggs, say θ_1 for the first alternative and θ_2 for the second one, with $\theta_1 > \theta_2$. This last condition leads to $\mathbb{E}(X) > \mathbb{E}(Y)$ and in such case it is more convenient to diversify if a large number of trips are going to be made. But for a single trip decision the condition $\theta_1 > \theta_2$ is not enough to prefer diversification unless $\theta_2(2 - \theta_2) < \theta_1$, since $\theta_2 < \theta_2(2 - \theta_2)$.

The main purpose of the present work is to show that the common belief that risk diversification is **always** better is more a **dogma**¹ rather than a general principle that has been proved, and that the correct view is to state that risk diversification may be better, as good as, or worse than lack thereof, depending on the risks involved and the dependence relationship among them.

2 Risk measures

Let X be a continuous random variable, with strictly increasing distribution function F_X , that represents an economic loss generated by certain events covered by insurance or related to investments. Without loss of generality we consider amounts of constant value over time (inflation indexed, for example). As a point estimation for a potential loss we may use the mean or the median. In the present work the median is preferred since it always exists for continuous random variables and it is robust, in contrast with the mean that may not exist or could be numerically unstable under heavy-tailed probability distributions. Using the quantile function (inverse of F_X) we calculate the median as $\mathbb{M}(X) = F_X^{-1}(\frac{1}{2})$ since $\mathbb{P}(X \leq \mathbb{M}(X)) = \frac{1}{2}$.

Definition 2.1. The *excess of loss* for a continuous loss random variable X is the random variable:

$$L := X - \mathbb{M}(X).$$

As suggested by McNeil *et al.*(2015) one way to interpret a *risk measure* is as the required additional risk capital $\varrho(L)$ to cover a loss in excess of what it was originally estimated. In the specialized literature on this subject there are many properties for risk measures that are considered as “desirable” or “reasonable”, though some concerns have been raised for some of them.

Definition 2.2. A risk measure ϱ is *monotone* if for any excess of loss random variables L_1 and L_2 such that $\mathbb{P}(L_1 \leq L_2) = 1$ we have that $\varrho(L_1) \leq \varrho(L_2)$.

McNeil *et al.*(2015) and several other authors consider monotonicity as a clearly desirable property since financial positions that involve higher risks under any circumstance should be covered by more risk capital. Positions such that $\varrho(L) \leq 0$ do not require additional capital.

Definition 2.3. A risk measure ϱ is *translation invariant* if for any excess of loss random variable L and any constant c we have that $\varrho(L + c) = \varrho(L) + c$.

This property is also considered as desirable by McNeil *et al.*(2015) and other authors under the following argument: the uncertainty associated to $L' := L + c$ totally depends on L since c is fixed, $\varrho(L)$ is the additional risk capital required to cover an excess of loss under L and therefore it would be enough to add the fixed amount c in order to cover for L' .

Definition 2.4. A risk measure ϱ is *subadditive* if for any excess of loss random variables L_1 and L_2 we have that $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$.

¹A system of principles or tenets; doctrine. A specific principle of a doctrine put forth, such as by a church. *Source:* WordReference Random House Learner’s Dictionary of American English © 2016.

This property cannot be considered as generally acceptable since there is some debate around it. One argument in favor is that diversification **always** reduces risk, which is more a **dogma** rather than something proved to be true under all circumstances. We may counter argue that for some risks there could be some sort of pernicious interaction that generates additional risk to the individual ones, so it may be also argued that it is better for a risk measure not to be subadditive, so that whenever it happens that $\varrho(L_1 + L_2) > \varrho(L_1) + \varrho(L_2)$ then it becomes clear that diversification is not convenient in such case.

Definition 2.5. A risk measure ϱ is *positively homogeneous* if for any excess of loss random variable L and any constant $\lambda > 0$ we have that $\varrho(\lambda L) = \lambda \varrho(L)$.

With regard to this property McNeil *et al.*(2015) and other authors mention that in case subadditivity has been accepted as reasonable then for any positive integer n it should be accepted that

$$\varrho(nL) = \varrho(L + \dots + L) \leq n\varrho(L) \quad (1)$$

and since there is no diversification “benefit” (because just a single risk source is involved) then the highest value would be attained in (1), that is equality. The same authors acknowledge there is some criticism about this property since for sufficiently large values of λ we should have $\varrho(\lambda L) > \lambda \varrho(L)$ to penalize for a high concentration of risk in a single source of it.

Definition 2.6. ϱ is a *coherent risk measure* if it satisfies Definitions 2.2 through 2.5.

The adjective “coherent” in this definition is somehow overbearing since it implicitly suggests that any risk measure that does not satisfy this definition would be incoherent despite the fact there is some debate and concerns about two of the four properties it requires. There are other additional properties that have been proposed in some contexts, see McNeil *et al.*(2015) or Denuit *et al.*(2005), but for the purpose of this article the above mentioned ones are enough.

3 Value at risk

As suggested by McNeil *et al.*(2015) we may interpret $\varrho(L)$ as the additional risk capital to cover for a potential excess of loss with L , but in practice such interpretation could be easily unachievable. Consider, for example, an insurance portfolio with certain face amounts for each issued policy. The only way to guarantee that the insurance company has enough resources to pay the claims under all possible scenarios would require the total reserve to be equal to the sum of all the face amounts in such portfolio.

In practice, specially under the *Basel Accords* and *Solvency II* frameworks, what is calculated is the amount of risk capital that has an *acceptable* high probability (but strictly less than 1) of covering an excess of loss that might face an insurance or financial institution. Who determines how much is “acceptable”? Typically the regulatory authority, but each company may decide to use probability levels even higher than the regulatory ones.

Definition 3.1. *Value at Risk* of level $0 < \alpha < 1$ for an excess of loss random variable L is a risk measure defined as

$$\text{VaR}_\alpha(L) := F_L^{-1}(\alpha)$$

where F_L^{-1} is the quantile function of L , that is the inverse of the probability distribution function of L .

In other words, a level α Value at Risk associated to a continuous random variable is the amount that such variable would not exceed with probability α . It should be noticed that the median is a Value at Risk of level $\alpha = \frac{1}{2}$.

Proposition 3.1. *VaR is a monotone, translation invariant, and positively homogeneous risk measure.*

Proof:

- a) Let X and Y be random variables such that $\mathbb{P}(X \leq Y) = 1$. Then for any value $x \in \mathbb{R}$:

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \leq x < Y) + \mathbb{P}(X \leq Y \leq x) \geq \mathbb{P}(\{X \leq Y\} \cap \{Y \leq x\}) = \mathbb{P}(Y \leq x),$$

that is $F_X(x) \geq F_Y(x)$. Let $x_\alpha := \text{VaR}_\alpha(X)$ and $y_\alpha := \text{VaR}_\alpha(Y)$. Then $\alpha = F_X(x_\alpha) \geq F_Y(x_\alpha)$ and since $\alpha = F_Y(y_\alpha)$ and distribution functions are non decreasing, necessarily $x_\alpha \leq y_\alpha$ and therefore $\text{VaR}_\alpha(X) \leq \text{VaR}_\alpha(Y)$.

- b) Let X be a continuous random variable with strictly increasing distribution function F_X and let $c \in \mathbb{R}$ be any given constant. Define the random variable $Y := X + c$, its probability distribution function is:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X + c \leq y) = \mathbb{P}(X \leq y - c) = F_X(y - c).$$

Let $x_\alpha := \text{VaR}_\alpha(X)$ and $y_\alpha := \text{VaR}_\alpha(Y)$. Then:

$$F_X(x_\alpha) = \alpha = F_Y(y_\alpha) = F_X(y_\alpha - c),$$

and since F_X is strictly increasing then $x_\alpha = y_\alpha - c$ which is equivalent to $\text{VaR}_\alpha(X) + c = \text{VaR}_\alpha(Y) = \text{VaR}_\alpha(X + c)$.

- c) Let X be a continuous random variable with strictly increasing distribution function F_X and let $\lambda > 0$ be a given constant. Define the random variable $Y := \lambda X$, its probability distribution function is:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\lambda X \leq y) = \mathbb{P}(X \leq y/\lambda) = F_X(y/\lambda).$$

Let $x_\alpha := \text{VaR}_\alpha(X)$ and $y_\alpha := \text{VaR}_\alpha(Y)$. Then:

$$F_X(x_\alpha) = \alpha = F_Y(y_\alpha) = F_X(y_\alpha/\lambda),$$

and since F_X is strictly increasing then $x_\alpha = y_\alpha/\lambda$ which is equivalent to $\lambda \text{VaR}_\alpha(X) = \text{VaR}_\alpha(Y) = \text{VaR}_\alpha(\lambda X)$. \square

It should be noticed that VaR is proved to be positively homogeneous without a subadditivity argument as in (1). In fact, VaR is not generally subadditive as it will become clear in a following section, but it will be also argued that this should not be considered as a disadvantage.

Example 3.1. Let X be a *Pareto* continuous random variable with parameters $\beta > 0$ y $\delta > 0$. Its probability density function is given by:

$$f_X(x | \beta, \delta) = \frac{\delta \beta^\delta}{x^{\delta+1}}, \quad x > \beta,$$

and therefore its probability distribution function:

$$F_X(t) = \int_{-\infty}^t f_X(x | \beta, \delta) dx = \delta \beta^\delta \int_{\beta}^t \frac{dx}{x^{\delta+1}} = 1 - \left(\frac{\beta}{t}\right)^\delta, \quad t > \beta.$$

The quantile function of X is the inverse of F_X , that is $F_X^{-1}(u) = \beta(1 - u)^{-1/\delta}$ for $0 < u < 1$ and consequently the median is $\mathbb{M}(X) = \text{VaR}_{1/2}(X) = F_X^{-1}(\frac{1}{2}) = 2^{1/\delta} \beta$. The level $\alpha > \frac{1}{2}$ VaR for the excess of loss $L = X - \mathbb{M}(X)$ is given by:

$$\text{VaR}_\alpha(L) = \text{VaR}_\alpha(X - \mathbb{M}(X)) = \text{VaR}_\alpha(X) - \mathbb{M}(X) = \beta[(1 - \alpha)^{-1/\delta} - 2^{1/\delta}].$$

Thus, with probability α the excess of loss will not exceed the amount $\text{VaR}_\alpha(L)$. Notice that if $\alpha \rightarrow 1-$ then $\text{VaR}_\alpha(L) \rightarrow +\infty$, which would require an infinite risk capital, something impossible in practice, and instead a value $\alpha < 1$ sufficiently close to 1 is arbitrarily set by the regulatory authority, for example $\alpha = 0.995$, though it is not clear how a particular value of α is considered “safe enough” in some sense.

As an additional comment for this last example, the mean for the *Pareto* model may not exist, it only does when $\delta > 1$ and even in such case $\mathbb{E}(X) = \beta\delta/(\delta - 1)$ which implies that for values of δ sufficiently close to 1 it is possible to have $\mathbb{E}(X) > \text{VaR}_\alpha(X)$ for any given value $\alpha < 1$ because $\lim_{\delta \rightarrow 1+} \mathbb{E}(X) = +\infty$. Since parameter δ controls tail heaviness of this probability distribution (lower δ values imply heavier right tail) this exemplifies a comment at the beginning of the previous section in the sense that it is better to use the median instead of the mean.

4 Loss aggregation

Consider n excess of loss random variables L_1, \dots, L_n where $L_i = X_i - \mathbb{M}(X_i)$ for $i \in \{1, \dots, n\}$ as in Definition 2.1. It is of interest to calculate VaR of the aggregation of such random variables:

$$L = L_1 + \dots + L_n = \sum_{i=1}^n X_i - \sum_{i=1}^n M(X_i) = S - c, \quad (2)$$

where the random variable $S := \sum_{i=1}^n X_i$ and the constant $c := \sum_{i=1}^n M(X_i)$. In this case we get $\text{VaR}_\alpha(L) = \text{VaR}_\alpha(S) - c$ so this last calculation essentially depends on obtaining or estimating the probability distribution function of S , that is F_S , because $\text{VaR}_\alpha(S) = F_S^{-1}(\alpha)$. Since S is a transformation of the n -dimensional random vector (X_1, \dots, X_n) it is necessary to know either the joint probability distribution function $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$ or its joint probability density function $f_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0$ such that

$$\mathbb{P}[(X_1, \dots, X_n) \in B] = \int_B \dots \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

A very popular probabilistic model is the *multivariate Normal* distribution, which undoubtedly has very nice mathematical properties that makes it very attractive for analysis and simplified calculations, but in practice it is usually inappropriate for the following reasons:

- All the univariate marginal distributions have to be *Normal*. Very often excess of loss random variables exhibit such a probabilistic behavior that are easily rejected by standard statistical normality tests, specially for heavier tails than the *Normal* distribution.
- The *multivariate Normal* is completely unable to consider *tail dependence* that very often is present among risks in finance and insurance, which consists in an important increase of the dependence degree under extreme values of the random variables involved.

These two flaws combined usually lead to a significant underestimation of the total aggregated risk. Instead, more flexible models have been explored, such as the ones built by means of *copula functions* which allow for any kind and distinct marginal univariate distributions and also account for tail dependence. Getting into the details of copula modeling is beyond the scope of the present article, the interested reader should refer to Nelsen (2006) for a book on basic copula theory, and the books by McNeil *et al.* (2015) and Denuit *et al.* (2005) for applications of copulas in finance and insurance risk modeling.

In two following sections, calculation of aggregated VaR will be considered in two extreme cases: perfect positive dependence (comonotonicity) and complete absence of dependence (that is, independence). For simplicity, but without loss of generality, it is considered the aggregation of two excess of loss random variables, that is $L = L_1 + L_2$ where $L_1 = X - \mathbb{M}(X)$ and $L_2 = Y - \mathbb{M}(Y)$, which is equivalent to $L = S - c$ with $S := X + Y$ and $c := \mathbb{M}(X) + \mathbb{M}(Y)$ and therefore $\text{VaR}_\alpha(L) = \text{VaR}_\alpha(S) - c$.

5 Comonotonicity

The following result comes from the works by Hoeffding (1940) and Fréchet (1951) and it is known as the *Fréchet-Hoeffding bounds* for joint probability distribution functions, which for simplicity is stated for the bivariate case:

Lemma 5.1. (*Fréchet-Hoeffding*) If (X, Y) is a random vector with joint probability distribution function $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ and marginal distribution functions $F_X(x) = \mathbb{P}(X \leq x)$ and $F_Y(y) = \mathbb{P}(Y \leq y)$ then:

$$H_*(x, y) := \max\{F_X(x) + F_Y(y) - 1, 0\} \leq F_{X,Y}(x, y) \leq \min\{F_X(x), F_Y(y)\} =: H^*(x, y),$$

where the lower bound H_* and the upper bound H^* are both joint distribution functions and therefore infimum and supremum for all bivariate joint distribution functions.

Definition 5.1. Two random variables X and Y are *comonotone* or *perfectly positively dependent* if there exists a strictly increasing function g such that $\mathbb{P}[Y = g(X)] = 1$.

Proof of the following lemma may be found in Nelsen (2006) as Theorem 2.5.4 and following comment thereof:

Lemma 5.2. (*Nelsen, 2006*) Let X and Y be continuous random variables with marginal distribution functions F_X and F_Y , respectively, and joint distribution function $F_{X,Y}$. Then X and Y are comonotone if and only if $F_{X,Y}$ is equal to the Fréchet-Hoeffding upper bound.

Now the main result for this section:

Theorem 5.1. If X and Y are continuous comonotone random variables then:

$$\text{VaR}_\alpha(X + Y) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y).$$

Proof:

Since X and Y are comonotone there exists a strictly increasing function g such that $\mathbb{P}[Y = g(X)] = 1$, hence the distribution function of Y may be expressed as:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}[g(X) \leq y] = \mathbb{P}[X \leq g^{-1}(y)] = F_X(g^{-1}(y)).$$

By Lemma 5.2 we get:

$$F_{X,Y}(x, y) = \min\{F_X(x), F_Y(y)\} = \min\{F_X(x), F_X(g^{-1}(y))\}.$$

Define $S := X + Y$, then its distribution function satisfies:

$$F_S(s) = \mathbb{P}(S \leq s) = \mathbb{P}(X + Y \leq s) = \mathbb{P}(X + g(X) \leq s) = \mathbb{P}(Y \leq s - X).$$

Since $\mathbb{P}[Y = g(X)] = 1$ then $F_{X,Y}$ is a singular distribution because all the probability is distributed along the curve $y = g(x)$ and therefore $F_S(s)$ is equal to the value of $F_{X,Y}$ at the intersection point (x_*, y_*) between the increasing curve $y = g(x)$ and the decreasing line $y = s - x$, for all $s \in \text{Ran } g$, which requires $g(x) = s - x$ and hence the intersection point is $(x_*, g(x_*))$ where x_* is the solution to the equation $x + g(x) = s$ which will be denoted as $x_* = h(s)$. Since g is strictly increasing so it is h which has inverse $h^{-1}(x) = x + g(x)$. Then:

$$F_S(s) = F_{X,Y}(x_*, g(x_*)) = \min\{F_X(x_*), F_X(g^{-1}(g(x_*)))\} = F_X(h(s)),$$

and consequently:

$$\begin{aligned} \text{VaR}_\alpha(X + Y) &= \text{VaR}_\alpha(S) = F_S^{-1}(\alpha) = h^{-1}(F_X^{-1}(\alpha)) \\ &= F_X^{-1}(\alpha) + g(F_X^{-1}(\alpha)) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) \quad \square \end{aligned}$$

Corollary 5.1. *If X and Y are continuous comonotone random variables, then for the excess of loss random variables $L_1 := X - \mathbb{M}(X)$ and $L_2 := Y - \mathbb{M}(Y)$ we have that:*

$$\text{VaR}_\alpha(L_1 + L_2) = \text{VaR}_\alpha(L_1) + \text{VaR}_\alpha(L_2).$$

Proof:

$$\begin{aligned} \text{VaR}_\alpha(L_1 + L_2) &= \text{VaR}_\alpha(X + Y - \mathbb{M}(X) - \mathbb{M}(Y)) = \text{VaR}_\alpha(X + Y) - \mathbb{M}(X) - \mathbb{M}(Y) \\ &= \text{VaR}_\alpha(X) - \mathbb{M}(X) + \text{VaR}_\alpha(Y) - \mathbb{M}(Y) = \text{VaR}_\alpha(L_1) + \text{VaR}_\alpha(L_2) \quad \square \end{aligned}$$

Example 5.1. Let X be a *Pareto* random variable with parameters $\beta = 1$ and $\delta > 0$ and define the random variable $Y := X^2$. Since $Y = g(X)$ with $g(x) = x^2$ a strictly increasing function on $\text{Ran } X =]1, +\infty[$ then X and Y are comonotone, with $\text{Ran } Y =]1, +\infty[$ also. Making use of the formulas in Example 3.1 we obtain:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) = 1 - \left(\frac{1}{y}\right)^{\delta/2}, \quad y > 1, \end{aligned}$$

which implies that Y is also a *Pareto* random variable but with parameters $\beta = 1$ and $\delta/2$, therefore:

$$\text{Var}_\alpha(X) = (1 - \alpha)^{-1/\delta}, \quad \text{Var}_\alpha(Y) = (1 - \alpha)^{-2/\delta}.$$

Now let $S := X + Y = X + X^2$ where $\text{Ran } S =]2, +\infty[$ and we get:

$$\begin{aligned} F_S(s) &= \mathbb{P}(S \leq s) = \mathbb{P}(X + X^2 \leq s) = \mathbb{P}(X \leq (\sqrt{1 + 4s} - 1)/2) \\ &= F_X((\sqrt{1 + 4s} - 1)/2) = 1 - (2/(\sqrt{1 + 4s} - 1))^\delta, \quad s > 2, \end{aligned}$$

from where we obtain for any $0 < \alpha < 1$ the following:

$$\text{VaR}_\alpha(X + Y) = \text{VaR}_\alpha(S) = F_S^{-1}(\alpha) = (1 - \alpha)^{-1/\delta} + (1 - \alpha)^{-2/\delta} = \text{Var}_\alpha(X) + \text{Var}_\alpha(Y),$$

as expected. \square

6 Independence

In contrast with the comonotonicity case where such property always implies that the VaR of the sum is equal to sum of the individual VaRs, under lack of dependence (independence) it is not possible to establish a general formula that relates the VaR for a sum of independent random variables to the individual VaRs, it will depend on each particular case, as it is shown in the following three examples:

Example 6.1. Let X and Y be independent and identically distributed *Pareto* random variables with parameters $\beta = 1$ and $\delta = 1$ such that the right tail of their distributions is heavy enough for non existence of a mean. Again applying formulas from Example 3.1 we get $\text{VaR}_\alpha(X) = (1 - \alpha)^{-1} = \text{VaR}_\alpha(Y)$ where $0 < \alpha < 1$ and by independence the joint density function for the random vector (X, Y) is the product of the marginal densities:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = \frac{1}{x^2y^2}, \quad x > 1, y > 1.$$

Let $S := X + Y$ then $\text{Ran } S =]2, +\infty[$ and its distribution function:

$$\begin{aligned} F_S(s) &= \mathbb{P}(S \leq s) = \mathbb{P}(X + Y \leq s) = \mathbb{P}(Y \leq s - X) = \iint_{y \leq s-x} f_{X,Y}(x, y) dx dy \\ &= \int_1^{s-1} x^{-2} \int_1^{s-x} y^{-2} dy dx = 1 - \frac{2}{s} - \frac{2}{s^2} \log(s-1), \quad s > 2. \end{aligned}$$

Let $s_* := \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) = 2/(1 - \alpha) > 2$. Then:

$$F_S(s_*) = \alpha - \frac{(1 - \alpha)^2}{2} \log \left(\frac{1 + \alpha}{1 - \alpha} \right) < \alpha,$$

which implies that for any $0 < \alpha < 1$:

$$\text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) = s_* < F_S^{-1}(\alpha) = \text{VaR}_\alpha(S) = \text{VaR}_\alpha(X + Y).$$

Despite total absence of dependence between the random variables the right tails of their distributions are heavy enough such that the diversification effect is definitely not convenient: the VaR of the sum is greater than the sum of the individual VaRs, in this particular case. \square

Example 6.2. Now let X and Y be independent and identically distributed *Normal* $(0, 1)$ random variables. Their distribution function is expressed as:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt.$$

The tails of this distribution are not as heavy as in the previous example, and it has finite mean and variance. Then the random variable $S := X + Y$ has *Normal* $(0, 2)$ distribution, which is the same as $\sqrt{2} X$ since a linear transformation of a *Normal* random variable is still *Normal* and $\mathbb{E}(\sqrt{2} X) = \sqrt{2} \mathbb{E}(X) = 0$ and $\mathbb{V}(\sqrt{2} X) = 2\mathbb{V}(X) = 2$. Therefore the distribution function of S may be expressed as:

$$F_S(s) = \mathbb{P}(S \leq s) = \mathbb{P}(\sqrt{2} X \leq s) = \mathbb{P}(X \leq s/\sqrt{2}) = \Phi(s/\sqrt{2}),$$

and its quantile function as $F_S^{-1}(u) = \sqrt{2} \Phi^{-1}(u)$, $0 < u < 1$. Consequently, for any $0 < \alpha < 1$:

$$\text{VaR}_\alpha(X + Y) = \text{VaR}_\alpha(S) = F_S^{-1}(\alpha) = \sqrt{2} \Phi^{-1}(\alpha) < 2\Phi^{-1}(\alpha) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y).$$

In contrast with the previous example, the VaR of this sum of random variables is strictly less than the sum of the individual VaRs, and therefore in this particular case diversification is clearly convenient. \square

Example 6.3. Lastly, let X and Y be independent and identically distributed *Exponential* random variables with parameter equal to 1. The right tail of this distribution is not as heavy as in Example 6.1 but certainly heavier than in Example 6.2, with finite mean and variance. Their marginal probability density function is $f(x) = e^{-x}$, $x > 0$, and the corresponding distribution function $F(x) = 1 - e^{-x}$, $x > 0$, hence $\text{VaR}_\alpha(X) = -\log(1 - \alpha) = \text{VaR}_\alpha(Y)$ where $0 < \alpha < 1$. By independence the joint density function of the random vector (X, Y) is the product of the marginal densities:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = e^{-(x+y)}, \quad x > 0, y > 0.$$

Let $S := X + Y$, then $\text{Ran } S =]0, +\infty[$ and its distribution function is:

$$\begin{aligned} F_S(s) &= \mathbb{P}(X + Y \leq s) = \iint_{y \leq s-x} f_{X,Y}(x, y) dx dy \\ &= \int_0^s e^{-x} \int_0^{s-x} e^{-y} dy dx = 1 - e^{-s}(1 + s), \quad s > 0. \end{aligned}$$

By the way, calculating the derivative of $F_S(s)$ we get $f_S(s) = se^{-s}$, $s > 0$, which is a density of a *Gamma* $(2, 1)$ random variable. Let $s_* := \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) = -2\log(1 - \alpha)$. Then:

$$g(\alpha) := F_S(s_*) = 1 - (1 - \alpha)^2(1 - 2\log(1 - \alpha)), \quad 0 < \alpha < 1.$$

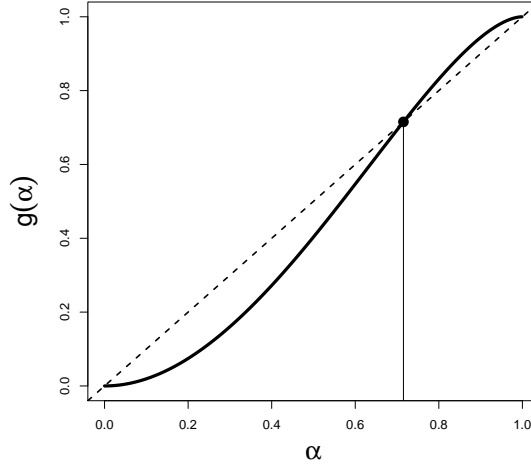


Figure 1: Graph of $g(\alpha) = F_S(-2 \log(1 - \alpha))$ in Example 6.3.

By numerical approximation it is obtained that $g(\alpha) = \alpha$ if and only if $\alpha \approx 0.7153319$, see Figure 1, $g(\alpha) < \alpha$ if $\alpha < 0.7153319$ and $g(\alpha) > \alpha$ if $\alpha > 0.7153319$, which implies that

$$\text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) \begin{cases} < \text{VaR}_\alpha(X + Y) & \text{if } \alpha < 0.7153319 \\ = \text{VaR}_\alpha(X + Y) & \text{if } \alpha \approx 0.7153319 \\ > \text{VaR}_\alpha(X + Y) & \text{if } \alpha > 0.7153319 \end{cases}$$

This is an example where diversification convenience depends on the desired α level for VaR, in contrast with the two previous examples. \square

7 Final remarks

The main conclusion in the present work is that diversification is **not always** convenient. As shown in the examples, risk diversification may result better, worse or equivalent to lack thereof, depending on the individual risks involved and the dependence relationship between them, and even on the desired risk level. In particular, as a consequence of Theorem 5.1, if two continuous random variables are comonotone then we can guarantee that the VaR is always equal to the sum of the individual VaRs. But for independent random variables everything may happen.

Moreover, it is argued that the fact VaR is not subadditive is more and advantage: in case the VaR of a sum is greater than the sum of individual VaRs we would be detecting a specially pernicious combination of risks on which is not convenient to diversify, while under “coherent” risk measures as in Definition 2.6 where subadditivity is always present it would not possible to detect such a harmful risk combination.

Bibliography

- Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R. (2005) *Actuarial Theory for Dependent Risks*. Wiley (Chichester).
- Fréchet, M. (1951) Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon* **14**, (Sect. A Ser. 3), 53–77.

- Hoeffding, W. (1940) Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* **5**, 179–223.
- McNeil, A.J., Frey, R., Embrechts, P. (2015) *Quantitative Risk Management*. Princeton University Press (New Jersey).
- Nelsen, R.B. (2006) *An Introduction to Copulas*. Springer (New York).

Valor en riesgo y el dogma de la diversificación

Arturo Erdely*

Facultad de Estudios Superiores Acatlán
Universidad Nacional Autónoma de México
arturo.erdely@comunidad.unam.mx

Resumen

Se analiza el principio de diversificación de riesgos y se demuestra que no siempre resulta mejor que no diversificar, pues esto depende de características individuales de los riesgos involucrados, así como de la relación de dependencia entre los mismos.

Palabras clave: valor en riesgo, agregación de pérdidas, comonotonicidad, diversificación.

1. Introducción

Un refrán popular sugiere que *no pongas todos los huevos en una misma canasta* y lleva implícito un principio (llamémoslo así momentáneamente) de *diversificación de riesgos* que más o menos tendría la siguiente “justificación”: Supongamos que necesitamos trasladar $2n$ huevos caminando de un punto A a un punto B y que tenemos acceso a dos alternativas, la primera, recurrir a una persona con una sola canasta con capacidad para la totalidad de los $2n$ huevos, y la segunda, recurrir a dos personas, cada una con una canasta con capacidad para n huevos, que de forma separada e independiente harían dicho traslado. El mencionado refrán sugiere que hay mayor riesgo en la primera alternativa, pues si la persona tropieza en el camino, se rompería la totalidad de los $2n$ huevos (pérdida total), y en cambio bajo la segunda alternativa si una de las personas tropieza sólo se perdería la mitad, y sería muy mala suerte que ambas tropezaran, en cuyo caso la pérdida agregada sería de todos modos la misma que si la persona de la primera alternativa tropezara.

Sea X una variable aleatoria que cuantifica (en número de huevos) la pérdida bajo la primera alternativa (una sola canasta), y sea Y la que cuantifica la pérdida bajo la segunda (dos canastas). Sea $0 < \theta < 1$ la probabilidad de que una persona tropiece y rompa los huevos al caminar del punto A al punto B. Entonces X, Y son variables aleatorias discretas tales que $\mathbb{P}(X \in \{0, 2n\}) = 1$ y $\mathbb{P}(Y \in \{0, n, 2n\}) = 1$, con probabilidades puntuales $\mathbb{P}(X = 2n) = \theta$, $\mathbb{P}(X = 0) = 1 - \theta$, $\mathbb{P}(Y = 2n) = \theta^2$, $\mathbb{P}(Y = n) = 2\theta(1 - \theta)$, $\mathbb{P}(Y = 0) = (1 - \theta)^2$. Si bien la probabilidad de tener la pérdida máxima $2n$ es mayor bajo la primera alternativa que bajo la segunda, notemos que la probabilidad de no tener pérdida alguna bajo la primera alternativa también es mayor que bajo la segunda. Más aún, notemos que

$$\mathbb{P}(Y > 0) = \theta^2 + 2\theta(1 - \theta) = \theta(2 - \theta) > \theta = \mathbb{P}(X > 0),$$

lo que quiere decir que es más probable que suframos algún tipo de pérdida (parcial o total) bajo la segunda alternativa que bajo la primera. Entonces... ¿es mejor poner todos los huevos en la misma canasta? Si el traslado de los huevos se va a realizar una sola vez, la respuesta sería afirmativa, pero si el mismo traslado va a realizarse un número muy grande de veces entonces deberíamos analizar la pérdida promedio de *largo*

*Sitio personal en internet <https://sites.google.com/site/arturoerdely>

plazo, que en este caso sería $\mathbb{E}(X) = 2n\theta$ para la primera alternativa, y $\mathbb{E}(Y) = 2n\theta^2 + 2n\theta(1 - \theta) = 2n\theta$ para la segunda alternativa, es decir, a largo plazo no habría diferencia entre ambas alternativas.

Entonces... ¿nunca conviene diversificar en dos canastas? Si la probabilidad de tropezar y romper los huevos es la misma con $2n$ huevos que con la mitad de ellos (que bien podría ser razonable hasta cierto valor de n) entonces el proverbio fallaría, pero quizás para un número n suficientemente grande deberíamos considerar probabilidades distintas de tropezar y romperlos, digamos θ_1 para la primera alternativa y θ_2 para la segunda, con $\theta_1 > \theta_2$. Esta última condición implicaría que $\mathbb{E}(X) > \mathbb{E}(Y)$ y en tal caso sí es conveniente diversificar, siempre que se realice un número muy grande de traslados. Pero si se realiza un solo traslado entonces la condición $\theta_1 > \theta_2$ no sería suficiente para que convenga diversificar, a menos que $\theta_2(2 - \theta_2) < \theta_1$, ya que $\theta_2 < \theta_2(2 - \theta_2)$.

En el presente trabajo se pretende demostrar que la creencia común en que diversificar riesgos **siempre** es mejor que no hacerlo, es más un **dogma**¹ que un principio universal científicamente comprobado, y que lo correcto es decir que diversificar riesgos puede resultar mejor, peor o igual, según el tipo de riesgos involucrados y la relación de dependencia entre ellos.

2. Medidas de riesgo

Sea X una variable aleatoria continua, con función de distribución de probabilidades estrictamente creciente F_X , que representa pérdida económica derivada de eventos contemplados en un contrato de seguro o inversión. Para efectos prácticos y sin pérdida de generalidad consideraremos cantidades monetarias a valor constante en el tiempo (por ejemplo, indexadas a la inflación). Como una estimación puntual de la pérdida puede utilizarse alguna medida de tendencia central como la media (esperanza) o la mediana, por ejemplo. Utilizaremos la mediana porque siempre existe para variables aleatorias continuas y es robusta, en contraste con la media que puede no existir o bien ser inestablemente grande bajo distribuciones de probabilidad con colas muy pesadas. La mediana se calcula por medio de la función de cuantiles (inversa de F_X), esto es $\mathbb{M}(X) = F_X^{-1}(\frac{1}{2})$ ya que $\mathbb{P}(X \leq \mathbb{M}(X)) = \frac{1}{2}$.

Definición 2.1. La *pérdida en exceso* a lo inicialmente estimado para una variable aleatoria continua X que representa pérdidas es también una variable aleatoria que se define:

$$L := X - \mathbb{M}(X).$$

Como se sugiere en McNeil *et al.*(2015) una de entre varias formas para interpretar una *medida de riesgo* es como la cantidad de capital adicional necesario para hacer frente a una pérdida en exceso que pudiera presentarse, misma que denotaremos $\varrho(L)$. Hay varias propiedades que en la literatura especializada se sugieren como “deseables” o “razonables” para cualquier medida de riesgo, algunas quizás son intuitivamente razonables, otras en ocasiones generan algunos cuestionamientos.

Definición 2.2. Una medida de riesgo ϱ es *monótona* si para cualesquiera variables aleatorias de pérdida en exceso L_1 y L_2 tales que $\mathbb{P}(L_1 \leq L_2) = 1$ se cumple $\varrho(L_1) \leq \varrho(L_2)$.

Respecto a esta propiedad, McNeil *et al.*(2015) y diversos autores consideran que la monotonidad es obviamente deseable ya que posiciones o transacciones financieras que involucren mayores pérdidas, bajo cualquier escenario, requieren mayor capital de riesgo. Posiciones tales que $\varrho(L) \leq 0$ no requieren capital adicional alguno.

¹Proposición tenida por cierta y como principio innegable. Conjunto de creencias de carácter indiscutible y obligado para los seguidores de cualquier religión. *Fuente:* Real Academia Española, <http://dle.rae.es/?id=E4earE8>

Definición 2.3. Una medida de riesgo ϱ es *invariante bajo traslación* si para cualquier variable aleatoria de pérdida en exceso L y una constante cualquiera c se cumple que $\varrho(L + c) = \varrho(L) + c$.

También esta propiedad es considerada como obviamente deseable por McNeil *et al.*(2015) y otros autores bajo un argumento como el siguiente: la incertidumbre asociada a $L' := L + c$ depende totalmente de L , no de c , y por tanto si $\varrho(L)$ es el capital adicional necesario para hacer frente a una pérdida en exceso que pudiera presentarse con L basta agregarle c para contar con el capital de riesgo necesario para cubrir lo análogo con L' .

Definición 2.4. Una medida de riesgo ϱ es *subaditiva* si para cualesquiera variables aleatorias de pérdida en exceso L_1 y L_2 se cumple que $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$.

La propiedad anterior no es considerada “obviamente razonable” ya que existen debates al respecto. Uno de los argumentos a favor es que diversificar **siempre** reduce el riesgo, lo cual es más un **dogma** que algo que haya sido formalmente demostrado que ocurre bajo cualquier circunstancia, además de que surge la duda sobre si cierto tipo de interacción entre dos o más posibles fuentes de pérdida pudieran generar pérdidas adicionales a las que de por sí y de forma individual pueden generar. Al contrario, es opinión de quien escribe que es mejor que una medida de riesgo no sea subaditiva, ya que si en un momento dado ocurre que $\varrho(L_1 + L_2) > \varrho(L_1) + \varrho(L_2)$ estaríamos detectando una combinación de riesgos especialmente perniciosa, y que por ello debiéramos evitar.

Definición 2.5. Una medida de riesgo ϱ es *positivamente homogénea* si para cualquier variable aleatoria de pérdida en exceso L y cualquier constante $\lambda > 0$ se cumple que $\varrho(\lambda L) = \lambda \varrho(L)$.

Respecto a esta propiedad McNeil *et al.*(2015) y otros autores comentan que en caso de que se haya aceptado como razonable la subaditividad entonces para cualquier entero positivo n se tendría que aceptar

$$\varrho(nL) = \varrho(L + \dots + L) \leq n\varrho(L) \quad (1)$$

y como no hay “beneficio” por diversificación al tratarse de la misma fuente de pérdida se alcanzaría el máximo valor posible en (1), es decir igualdad. Los mismos autores reconocen la crítica que existe respecto a esta propiedad ya que hay quienes opinan que en ciertos contextos y con valores suficientemente grandes de λ debería cumplirse que $\varrho(\lambda L) > \lambda \varrho(L)$ para penalizar una elevada concentración del riesgo.

Definición 2.6. ϱ es una *medida coherente de riesgo* si cumple con las Definiciones 2.2 a 2.5.

El adjetivo “coherente” en la definición anterior resulta un tanto chocante, ya que implícitamente califica de incoherente a cualquier medida de riesgo que no la cumpla, a pesar de que existen cuestionamientos razonables sobre dos de las cuatro propiedades que exige. Existen propiedades adicionales que también se han propuesto en diversos contextos, véase el ya multicitado libro de McNeil *et al.*(2015) o bien Denuit *et al.*(2005), pero para el alcance que se pretende en el presente artículo lo anterior es suficiente.

3. Valor en riesgo

Si bien podemos interpretar a $\varrho(L)$ como el capital de riesgo necesario para hacer frente a una pérdida en exceso que pudiera presentarse con L , tal y como lo proponen McNeil *et al.*(2015), en la práctica dicha interpretación podría fácilmente resultar inviable. Pensemos, por ejemplo, en una cartera de pólizas de seguro con determinadas sumas aseguradas. La única forma de garantizar que se cuenta con suficientes recursos económicos para hacer frente a todas las posibles reclamaciones es que el total de reservas de la compañía de seguros fuese exactamente igual a la suma de todas las sumas aseguradas en dicha cartera.

Lo que usualmente se busca en la práctica, especialmente bajo los esquemas de los *Acuerdos de Basilea y Solvencia II*², es contar con un capital de riesgo que tenga una probabilidad “aceptablemente alta” (pero estrictamente menor que 1) de cubrir el total de pérdida en exceso que pudiera enfrentar una entidad financiera o de seguros. ¿Quién determina cuánto es “aceptablemente alto”? Típicamente la autoridad reguladora del sector que corresponda, aunque adicionalmente cada entidad financiera o de seguros está en libertad de aplicar niveles de probabilidad aún mayores a los que como mínimo solicite el regulador.

Definición 3.1. Se denomina *valor en riesgo* de nivel $0 < \alpha < 1$ para una pérdida en exceso L a la medida de riesgo denotada y definida como

$$\text{VaR}_\alpha(L) := F_L^{-1}(\alpha)$$

donde las siglas VaR corresponden en idioma inglés a *Value at Risk* y F_L^{-1} es la función de cuantiles de la variable aleatoria continua L , esto es, la función inversa de la función de distribución de probabilidades de L .

Dicho de otra forma, el valor en riesgo de nivel α asociado a una variable aleatoria continua es una cantidad que dicha variable no excederá con probabilidad α . Nótese que la mediana es un valor en riesgo de nivel $\frac{1}{2}$.

Proposición 3.1. *La medida de riesgo VaR es monótona, invariante bajo traslación y positivamente homogénea.*

Demostración:

- a) Sean X, Y variables aleatorias continuas tales que $\mathbb{P}(X \leq Y) = 1$. Entonces para todo valor $x \in \mathbb{R}$:

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \leq x < Y) + \mathbb{P}(X \leq Y \leq x) \geq \mathbb{P}(\{X \leq Y\} \cap \{Y \leq x\}) = \mathbb{P}(Y \leq x),$$

es decir $F_X(x) \geq F_Y(x)$. Sean $x_\alpha := \text{VaR}_\alpha(X)$, $y_\alpha := \text{VaR}_\alpha(Y)$. Entonces $\alpha = F_X(x_\alpha) \geq F_Y(x_\alpha)$ y como también $\alpha = F_Y(y_\alpha)$ y las funciones de distribución son monótonas crecientes, necesariamente $x_\alpha \leq y_\alpha$ y por lo tanto $\text{VaR}_\alpha(X) \leq \text{VaR}_\alpha(Y)$.

- b) Sea X una variable aleatoria continua con función de distribución de probabilidades F_X estrictamente creciente y sea $c \in \mathbb{R}$ una constante cualquiera. Definiendo una variable aleatoria $Y := X + c$ tenemos que su función de distribución de probabilidades resulta ser:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X + c \leq y) = \mathbb{P}(X \leq y - c) = F_X(y - c).$$

Sean $x_\alpha := \text{VaR}_\alpha(X)$, $y_\alpha := \text{VaR}_\alpha(Y)$. Entonces:

$$F_X(x_\alpha) = \alpha = F_Y(y_\alpha) = F_X(y_\alpha - c),$$

y como F_X es estrictamente creciente entonces necesariamente $x_\alpha = y_\alpha - c$ lo cual equivale a $\text{VaR}_\alpha(X) + c = \text{VaR}_\alpha(Y) = \text{VaR}_\alpha(X + c)$.

- c) Sea X una variable aleatoria continua con función de distribución de probabilidades F_X estrictamente creciente y sea $\lambda > 0$ una constante. Definiendo una variable aleatoria $Y := \lambda X$ tenemos que su función de distribución de probabilidades resulta ser:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\lambda X \leq y) = \mathbb{P}(X \leq y/\lambda) = F_X(y/\lambda).$$

Sean $x_\alpha := \text{VaR}_\alpha(X)$, $y_\alpha := \text{VaR}_\alpha(Y)$. Entonces:

$$F_X(x_\alpha) = \alpha = F_Y(y_\alpha) = F_X(y_\alpha/\lambda),$$

y como F_X es estrictamente creciente entonces necesariamente $x_\alpha = y_\alpha/\lambda$ lo cual a su vez equivale a $\lambda \text{VaR}_\alpha(X) = \text{VaR}_\alpha(Y) = \text{VaR}_\alpha(\lambda X)$. \square

²Consúltese, por ejemplo, *Un marco global para la evaluación de la solvencia del asegurador*, Informe del Grupo de Trabajo para la Evaluación de la Solvencia del Asegurador de la Asociación Actuarial Internacional (2009).

Nótese en particular que la medida de riesgo VaR cumple con ser positivamente homogénea sin necesidad de aceptar subaditividad como en (1). De hecho VaR no es en general una medida de riesgo subaditiva como se mostrará más adelante, pero veremos también que esto no es necesariamente una desventaja.

Ejemplo 3.1. Sea X una variable aleatoria continua *Pareto* con parámetros $\beta > 0$ y $\delta > 0$. Su función de densidad de probabilidades es:

$$f_X(x | \beta, \delta) = \frac{\delta \beta^\delta}{x^{\delta+1}}, \quad x > \beta,$$

y por tanto su función de distribución de probabilidades resulta ser:

$$F_X(t) = \int_{-\infty}^t f_X(x | \beta, \delta) dx = \delta \beta^\delta \int_{\beta}^t \frac{dx}{x^{\delta+1}} = 1 - \left(\frac{\beta}{t}\right)^\delta, \quad t > \beta.$$

La función de cuantiles de X es la inversa de F_X , esto es $F_X^{-1}(u) = \beta(1-u)^{-1/\delta}$ para $0 < u < 1$ y a partir de esto último la mediana resulta ser $\mathbb{M}(X) = \text{VaR}_{1/2}(X) = F_X^{-1}(\frac{1}{2}) = 2^{1/\delta} \beta$. Calculemos ahora el valor en riesgo de nivel $\alpha > \frac{1}{2}$ para la pérdida en exceso $L = X - \mathbb{M}(X)$:

$$\text{VaR}_\alpha(L) = \text{VaR}_\alpha(X - \mathbb{M}(X)) = \text{VaR}_\alpha(X) - \mathbb{M}(X) = \beta[(1-\alpha)^{-1/\delta} - 2^{1/\delta}].$$

Así, con probabilidad α la pérdida en exceso no excederá la cantidad $\text{VaR}_\alpha(L)$. Nótese que si $\alpha \rightarrow 1-$ entonces $\text{VaR}_\alpha(L) \rightarrow +\infty$, lo cual requeriría un capital de riesgo infinito, algo imposible en la práctica, y es por ello que normalmente se elige un valor $\alpha < 1$ pero lo suficientemente cercano a 1 como para que la autoridad reguladora se sienta aceptablemente tranquila, por ejemplo $\alpha = 0.995$, aunque no queda claro cómo especificar un valor α que refleje un cierto nivel abstracto de “tranquilidad”.

Como comentario adicional respecto al ejemplo anterior, la esperanza para el modelo *Pareto* no siempre existe, solo existe cuando $\delta > 1$ y aún en tal caso se tiene que $\mathbb{E}(X) = \beta\delta/(\delta-1)$ lo cual implica que con valores de δ suficientemente cercanos a 1 por la derecha es posible que $\mathbb{E}(X)$ sea mayor que $\text{VaR}_\alpha(X)$ para cualquier valor dado $\alpha < 1$ ya que $\lim_{\delta \rightarrow 1+} \mathbb{E}(X) = +\infty$. Como el parámetro δ controla el grado de pesadez en la cola de la distribución de probabilidad en este modelo (a menor valor de δ mayor pesadez) esto ilustra el comentario al inicio de la sección anterior en el sentido de que es más conveniente utilizar la mediana en lugar de la media.

4. Agregación de pérdidas

Consideremos ahora n variables aleatorias de pérdida en exceso L_1, \dots, L_n en donde cada una se expresa como en la Definición 2.1, es decir $L_i = X_i - \mathbb{M}(X_i)$ para $i \in \{1, \dots, n\}$. Supongamos que es de interés calcular el valor en riesgo para la *agregación* o suma de dichas variables:

$$L = L_1 + \dots + L_n = \sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{M}(X_i) = S - c, \quad (2)$$

en donde se definen la variable aleatoria $S := \sum_{i=1}^n X_i$ y la constante $c := \sum_{i=1}^n \mathbb{M}(X_i)$. En este caso tendríamos que $\text{VaR}_\alpha(L) = \text{VaR}_\alpha(S) - c$ por lo que dicho cálculo depende esencialmente de poder obtener o estimar la función de distribución de probabilidades de S , es decir F_S , ya que $\text{VaR}_\alpha(S) = F_S^{-1}(\alpha)$. Como S es una transformación del vector aleatorio n -dimensional (X_1, \dots, X_n) entonces para la obtención o estimación de F_S se requiere una distribución de probabilidades conjunta que capture adecuadamente las dependencias entre las variables aleatorias que integran dicho vector aleatorio, ya sea una función de distribución conjunta de probabilidades $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$ o bien una función de densidad de probabilidades conjunta $f_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0$ tal que

$$\mathbb{P}[(X_1, \dots, X_n) \in B] = \int_B \dots \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Un modelo probabilístico muy popular es la distribución de probabilidad *Normal multivariada*, que si bien tiene propiedades matemáticas que la hacen muy atractiva para el análisis y simplificar cálculos, resulta con mucha frecuencia un modelo inapropiado por las siguientes razones:

- Las distribuciones marginales univariadas deben todas tener distribución *Normal*. Con frecuencia las variables de pérdida exhiben distribuciones de probabilidad que son rechazadas por pruebas estadísticas de normalidad, típicamente por tener colas más pesadas.
- La distribución *Normal multivariada* es incapaz de incorporar *dependencia en las colas* (*tail dependence* en idioma inglés), una característica que con frecuencia se observa entre variables asociadas a riesgos en seguros y finanzas y que consiste en un incremento importante en el grado de dependencia bajo valores extremos de las variables involucradas.

Estas dos deficiencias conducen a una subestimación del riesgo total agregado, lo que ha motivado la búsqueda de modelos probabilísticos más flexibles, como los que se pueden construir por medio de *funciones cópula*, mismos que permiten utilizar distribuciones marginales univariadas de cualquier tipo y distintas para cada variable involucrada, y además incorporar dependencia en las colas (*tail dependence*). Entrar al detalle de esto sería motivo de otro artículo, por lo pronto simplemente se hace referencia a Nelsen (2006) cuyo libro es considerado fundamental para la comprensión de la teoría básica de funciones cópula, y nuevamente los libros de McNeil *et al.*(2015) o bien Denuit *et al.*(2005) para su aplicación en finanzas y seguros. Para una breve introducción a funciones cópula puede revisarse Erdely (2009).

En las dos secciones siguientes se analizará el cálculo del valor en riesgo de una agregación de variables aleatorias en dos casos extremos: dependencia positiva perfecta (comonotonicidad) y ausencia total de dependencia (independencia). Por simplicidad pero sin pérdida de generalidad analizaremos el caso de la agregación de dos variables de pérdida en exceso, esto es $L = L_1 + L_2$ donde $L_1 = X - \mathbb{M}(X)$ y $L_2 = Y - \mathbb{M}(Y)$, que podemos expresar como $L = S - c$ donde $S := X + Y$ y $c := \mathbb{M}(X) + \mathbb{M}(Y)$ y por tanto $\text{VaR}_\alpha(L) = \text{VaR}_\alpha(S) - c$.

5. Comonotonicidad

El siguiente es un resultado consecuencia de los trabajos de Hoeffding (1940) y Fréchet (1951) conocido como *cotas de Fréchet-Hoeffding* para funciones de distribución de probabilidad conjunta, que por simplicidad aquí lo restringimos al caso bivariado:

Lema 5.1. (*Fréchet-Hoeffding*) Si (X, Y) es un vector aleatorio con función de distribución conjunta $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ y funciones de distribución marginales $F_X(x) = \mathbb{P}(X \leq x)$ y $F_Y(y) = \mathbb{P}(Y \leq y)$ entonces:

$$H_*(x, y) := \max\{F_X(x) + F_Y(y) - 1, 0\} \leq F_{X,Y}(x, y) \leq \min\{F_X(x), F_Y(y)\} =: H^*(x, y),$$

en donde la cota inferior H_* y la cota superior H^* son ambas funciones de distribución conjunta y por tanto constituyen ínfimo y supremo de todas las funciones de distribución conjunta bivariadas.

Definición 5.1. Se dice que dos variables aleatorias X, Y son *comonótonas* o bien que tienen *dependencia positiva perfecta* si existe una función g estrictamente creciente tal $\mathbb{P}[Y = g(X)] = 1$.

La demostración del siguiente lema puede consultarse en Nelsen (2006) como Teorema 2.5.4 y comentario posterior, mismo que resulta necesario para el teorema principal de esta sección:

Lema 5.2. (Nelsen, 2006) Sean X, Y variables aleatorias continuas con funciones de distribución marginal F_X y F_Y , respectivamente, y función de distribución conjunta $F_{X,Y}$. Entonces X, Y son comonótonas si y sólo si $F_{X,Y}$ es igual a la cota superior de Fréchet-Hoeffding.

A continuación, el resultado principal de esta sección:

Teorema 5.1. Si X, Y son variables aleatorias continuas y comonótonas entonces:

$$\text{VaR}_\alpha(X + Y) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y).$$

Demostración:

Como X, Y son comonótonas entonces existe una función g estrictamente creciente tal que $\mathbb{P}[Y = g(X)] = 1$, por lo que la función de distribución de Y puede expresarse como:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}[g(X) \leq y] = \mathbb{P}[X \leq g^{-1}(y)] = F_X(g^{-1}(y)).$$

Aplicando el Lema 5.2 tenemos que:

$$F_{X,Y}(x, y) = \min\{F_X(x), F_Y(y)\} = \min\{F_X(x), F_X(g^{-1}(y))\}.$$

Si se define $S := X + Y$ entonces su función de distribución de probabilidades satisface lo siguiente:

$$F_S(s) = \mathbb{P}(S \leq s) = \mathbb{P}(X + Y \leq s) = \mathbb{P}(X + g(X) \leq s) = \mathbb{P}(Y \leq s - X).$$

Como $\mathbb{P}[Y = g(X)] = 1$ entonces $F_{X,Y}$ es una distribución singular ya que toda la probabilidad se encuentra concentrada sobre la curva $y = g(x)$ y por lo tanto $F_S(s)$ es igual al valor acumulado por $F_{X,Y}$ en el punto de intersección (x_*, y_*) de la curva creciente $y = g(x)$ con la recta decreciente $y = s - x$, para todo valor $s \in \text{Ran } g$, lo cual requiere que $g(x) = s - x$ y por tanto el punto de intersección es $(x_*, g(x_*))$ donde x_* es la solución de la ecuación $x + g(x) = s$ que denotaremos $x_* = h(s)$. Como g es estrictamente creciente entonces h también lo es y tiene inversa $h^{-1}(x) = x + g(x)$. Entonces:

$$F_S(s) = F_{X,Y}(x_*, g(x_*)) = \min\{F_X(x_*), F_X(g^{-1}(g(x_*)))\} = F_X(h(s)),$$

y por lo tanto:

$$\begin{aligned} \text{VaR}_\alpha(X + Y) &= \text{VaR}_\alpha(S) = F_S^{-1}(\alpha) = h^{-1}(F_X^{-1}(\alpha)) \\ &= F_X^{-1}(\alpha) + g(F_X^{-1}(\alpha)) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) \quad \square \end{aligned}$$

Corolario 5.1. Si X, Y son variables aleatorias continuas comonótonas que representan pérdidas, entonces para las variables aleatorias de exceso de pérdida $L_1 := X - \mathbb{M}(X)$ y $L_2 := Y - \mathbb{M}(Y)$ se cumple que:

$$\text{VaR}_\alpha(L_1 + L_2) = \text{VaR}_\alpha(L_1) + \text{VaR}_\alpha(L_2).$$

Demostración:

$$\begin{aligned} \text{VaR}_\alpha(L_1 + L_2) &= \text{VaR}_\alpha(X + Y - \mathbb{M}(X) - \mathbb{M}(Y)) = \text{VaR}_\alpha(X + Y) - \mathbb{M}(X) - \mathbb{M}(Y) \\ &= \text{VaR}_\alpha(X) - \mathbb{M}(X) + \text{VaR}_\alpha(Y) - \mathbb{M}(Y) = \text{VaR}_\alpha(L_1) + \text{VaR}_\alpha(L_2) \quad \square \end{aligned}$$

Ejemplo 5.1. Sea X una variable aleatoria *Pareto* con parámetros $\beta = 1$ y $\delta > 0$ y defínase la variable aleatoria $Y := X^2$. Como $Y = g(X)$ con $g(x) = x^2$ una función estrictamente creciente sobre $\text{Ran } X =]1, +\infty[$ entonces X, Y son comonótonas, con $\text{Ran } Y =]1, +\infty[$ también. Aprovechando las fórmulas del Ejemplo 3.1 obtenemos:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) = 1 - \left(\frac{1}{y}\right)^{\delta/2}, \quad y > 1, \end{aligned}$$

lo cual implica que Y tiene distribución de probabilidad *Pareto* pero con parámetros $\beta = 1$ y $\delta/2$, y por tanto:

$$\text{Var}_\alpha(X) = (1 - \alpha)^{-1/\delta}, \quad \text{Var}_\alpha(Y) = (1 - \alpha)^{-2/\delta}.$$

Ahora se define la variable aleatoria $S := X + Y = X + X^2$ donde $\text{Ran } S =]2, +\infty[$ y se obtiene:

$$\begin{aligned} F_S(s) &= \mathbb{P}(S \leq s) = \mathbb{P}(X + X^2 \leq s) = \mathbb{P}(X \leq (\sqrt{1 + 4s} - 1)/2) \\ &= F_X((\sqrt{1 + 4s} - 1)/2) = 1 - (2/(\sqrt{1 + 4s} - 1))^\delta, \quad s > 2, \end{aligned}$$

de donde para todo valor $0 < \alpha < 1$ se obtiene:

$$\text{VaR}_\alpha(X + Y) = \text{VaR}_\alpha(S) = F_S^{-1}(\alpha) = (1 - \alpha)^{-1/\delta} + (1 - \alpha)^{-2/\delta} = \text{Var}_\alpha(X) + \text{Var}_\alpha(Y),$$

como era de esperarse. \square

6. Independencia

En contraste con el caso de comonotonicidad donde dicha característica implica necesariamente que el valor en riesgo de una suma de variables aleatorias es igual a la suma de los valores en riesgo individuales, bajo ausencia total de dependencia (es decir, independencia) no es posible establecer de forma general qué tipo de relación existirá entre el valor en riesgo de dicha suma y la suma de los valores en riesgo individuales, dependerá de cada caso particular, y para demostrarlo bastará con los siguientes tres ejemplos:

Ejemplo 6.1. Sean X, Y variables aleatorias independientes e idénticamente distribuidas *Pareto* con parámetros $\beta = 1$ y $\delta = 1$, que con dichos valores de los parámetros la cola derecha de la distribución de probabilidad es suficientemente pesada como para que esperanza y varianza no existan. Aprovechando las fórmulas del Ejemplo 3.1 obtenemos que $\text{VaR}_\alpha(X) = (1 - \alpha)^{-1} = \text{VaR}_\alpha(Y)$ donde $0 < \alpha < 1$ y además, por independencia, la función de densidad conjunta del vector aleatorio (X, Y) es igual al producto de las densidades marginales, esto es:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = \frac{1}{x^2y^2}, \quad x > 1, y > 1.$$

Si se define la variable aleatoria $S := X + Y$ entonces $\text{Ran } S =]2, +\infty[$ y su función de distribución de probabilidades es:

$$\begin{aligned} F_S(s) &= \mathbb{P}(S \leq s) = \mathbb{P}(X + Y \leq s) = \mathbb{P}(Y \leq s - X) = \iint_{y \leq s-x} f_{X,Y}(x, y) dx dy \\ &= \int_1^{s-1} x^{-2} \int_1^{s-x} y^{-2} dy dx = 1 - \frac{2}{s} - \frac{2}{s^2} \log(s - 1), \quad s > 2. \end{aligned}$$

Sea $s_* := \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) = 2/(1 - \alpha) > 2$. Entonces:

$$F_S(s_*) = \alpha - \frac{(1 - \alpha)^2}{2} \log\left(\frac{1 + \alpha}{1 - \alpha}\right) < \alpha,$$

lo cual implica que para todo valor $0 < \alpha < 1$:

$$\text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) = s_* < F_S^{-1}(\alpha) = \text{VaR}_\alpha(S) = \text{VaR}_\alpha(X + Y).$$

A pesar de la ausencia de dependencia alguna entre las variables aleatorias involucradas, las colas de sus distribuciones individuales son suficientemente pesadas como para que el efecto de diversificación resulte inconveniente, pues el valor en riesgo de su suma resulta mayor que la suma de los valores en riesgo individuales. \square

Ejemplo 6.2. Ahora sean X, Y variables aleatorias independientes e idénticamente distribuidas *Normal* $(0, 1)$ cuya función de distribución de probabilidades se obtiene mediante:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt.$$

Las colas de esta distribución de probabilidades no son tan pesadas como las del ejemplo anterior y tiene esperanza y varianza finitas. Es un conocido y elemental resultado de probabilidad que la variable aleatoria $S := X + Y$ tiene distribución de probabilidad *Normal* $(0, 2)$ y por ello S tiene la misma distribución de probabilidad que $\sqrt{2}X$ ya que cualquier transformación lineal de una variable aleatoria *Normal* sigue siendo *Normal* y además $\mathbb{E}(\sqrt{2}X) = \sqrt{2}\mathbb{E}(X) = 0$ y $\mathbb{V}(\sqrt{2}X) = 2\mathbb{V}(X) = 2$. Por lo anterior, la función de distribución de probabilidades de S puede expresarse de la siguiente manera:

$$F_S(s) = \mathbb{P}(S \leq s) = \mathbb{P}(\sqrt{2}X \leq s) = \mathbb{P}(X \leq s/\sqrt{2}) = \Phi(s/\sqrt{2}),$$

y su función de cuantiles mediante $F_S^{-1}(u) = \sqrt{2}\Phi^{-1}(u)$, $0 < u < 1$, por lo que para todo valor $0 < \alpha < 1$:

$$\text{VaR}_\alpha(X + Y) = \text{VaR}_\alpha(S) = F_S^{-1}(\alpha) = \sqrt{2}\Phi^{-1}(\alpha) < 2\Phi^{-1}(\alpha) = \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y).$$

En contraste con el ejemplo anterior, el valor en riesgo de esta suma de variables aleatorias independientes es menor que la suma de los valores en riesgo individuales, y por tanto en este caso particular la diversificación resulta conveniente. \square

Ejemplo 6.3. Finalmente sean X, Y variables aleatorias independientes e idénticamente distribuidas *Exponencial* estándar (parámetro igual a 1). La cola derecha de esta distribución de probabilidad no es tan pesada como la del Ejemplo 6.1 pero sí más pesada que en el Ejemplo 6.2, y tiene esperanza y varianza finitas. Su función de densidad de probabilidades es $f(x) = e^{-x}$, $x > 0$, su función de distribución $F(x) = 1 - e^{-x}$, $x > 0$ y $\text{VaR}_\alpha(X) = -\log(1 - \alpha) = \text{VaR}_\alpha(Y)$ donde $0 < \alpha < 1$. Además, por independencia, la función de densidad conjunta del vector aleatorio (X, Y) es igual al producto de las densidades marginales, esto es:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = e^{-(x+y)}, \quad x > 0, y > 0.$$

Si se define la variable aleatoria $S := X + Y$ entonces $\text{Ran } S =]0, +\infty[$ y su función de distribución de probabilidades es:

$$\begin{aligned} F_S(s) &= \mathbb{P}(X + Y \leq s) = \iint_{y \leq s-x} f_{X,Y}(x, y) dx dy \\ &= \int_0^s e^{-x} \int_0^{s-x} e^{-y} dy dx = 1 - e^{-s}(1 + s), \quad s > 0. \end{aligned}$$

Y de hecho derivando $F_S(s)$ la función de densidad resultante es $f_S(s) = se^{-s}$, $s > 0$, que corresponde a una distribución de probabilidad *Gamma* $(2, 1)$. Sea $s_* := \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) = -2\log(1 - \alpha)$. Entonces:

$$g(\alpha) := F_S(s_*) = 1 - (1 - \alpha)^2(1 - 2\log(1 - \alpha)), \quad 0 < \alpha < 1.$$

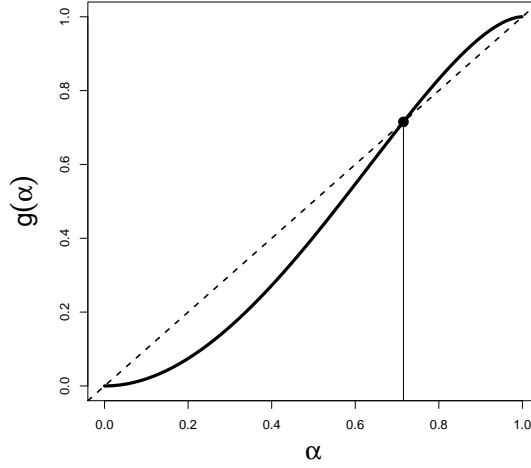


Figura 1: Gráfica de $g(\alpha) = F_S(-2 \log(1 - \alpha))$ en el Ejemplo 6.3.

Por aproximación numérica es verificable que $g(\alpha) = \alpha$ si y sólo si $\alpha \approx 0.7153319$, ver Figura 1, que $g(\alpha) < \alpha$ si $\alpha < 0.7153319$ y $g(\alpha) > \alpha$ si $\alpha > 0.7153319$, lo cual implica que

$$\text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y) \begin{cases} < \text{VaR}_\alpha(X + Y) & \text{si } \alpha < 0.7153319 \\ = \text{VaR}_\alpha(X + Y) & \text{si } \alpha \approx 0.7153319 \\ > \text{VaR}_\alpha(X + Y) & \text{si } \alpha > 0.7153319 \end{cases}$$

Aquí estamos ante un ejemplo en que la conveniencia o no de diversificar depende del nivel α de valor en riesgo deseado, situación que no ocurrió en los dos ejemplos anteriores. \square

7. Conclusiones

La principal conclusión del presente análisis es que resulta falsa la idea de que diversificar riesgos siempre es mejor que no hacerlo. Como se pudo ilustrar en diversos ejemplos, diversificar riesgos puede resultar mejor, peor o igual que no hacerlo, dependiendo de las características individuales de los riesgos involucrados, de la relación de dependencia entre ellos, e incluso hasta del nivel de riesgo deseado. En particular, como consecuencia del Teorema 5.1, si dos variables aleatorias continuas son comonótonas entonces sí es posible afirmar que en general el valor en riesgo de la suma es igual a la suma de los valores en riesgo individuales. Pero en el caso de variables aleatorias independendientes puede resultar mejor, peor o igual diversificar, dependiendo de las distribuciones de probabilidad de los riesgos involucrados.

Por lo anterior, se argumenta también que la no subaditividad del VaR es más una ventaja que desventaja, pues en aquellos casos en que el VaR de una agregación de riesgos resulte mayor que la suma de los VaR individuales estaríamos obteniendo inmediatamente información sobre cierto tipo de interacción entre riesgos que conviene evitar. En cambio, con medidas “coherentes” de riesgo como en la Definición 2.6 donde la subaditividad está garantizada, no tendríamos aviso alguno sobre interacciones de riesgo perniciosas.

Bibliografía

Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R. (2005) *Actuarial Theory for Dependent Risks*. Wiley (Chichester).

- Erdely, A. (2009) Cópulas y dependencia de variables aleatorias: una introducción. *Miscelánea matemática* **48**, 7–28.
- Fréchet, M. (1951) Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon* **14**, (Sect. A Ser. 3), 53–77.
- Hoeffding, W. (1940) Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* **5**, 179–223.
- McNeil, A.J., Frey, R., Embrechts, P. (2015) *Quantitative Risk Management*. Princeton University Press (New Jersey).
- Nelsen, R.B. (2006) *An Introduction to Copulas*. Springer (New York).